

Vorwort

Im Studium der Linguistik, der Fremdsprachenfächer und der Sprachlehr- und -lernforschung wird erwartet, dass Sie empirische Studien lesen, verstehen und in Bezug zu einander setzen können, um an Seminardiskussionen teilzunehmen und schriftliche Arbeiten zu verfassen. Dabei ist es manchmal schwer, sich unter den vielen Arten von wissenschaftlichen Studien zurechtzufinden.

Dieses Buch handelt von solchen empirischen Untersuchungen, die quantitativ vorgehen (also mit Zahlen arbeiten). Sie erhalten einen Überblick darüber, wann empirisches Arbeiten sinnvoll und angemessen ist, welche Verfahren es in der Linguistik und Sprachlehrforschung gibt und Hinweise darauf, wie man die Qualität von empirischen Arbeiten beurteilen kann. Das soll Ihnen helfen, sich in Ihren Hausarbeiten und Referaten auf wirklich aussagekräftige empirische Untersuchungen zu stützen und angreifbare Untersuchungen durch gute Argumentationen zu widerlegen.

Um den Text möglichst gut lesbar zu gestalten, haben wir in den Fällen, in denen keine gebräuchlichen geschlechtsneutralen Personenbezeichnungen existieren, mit etwas Bauchschmerzen die maskuline Form als Bezeichnung für alle Geschlechter gewählt.

Möglicherweise finden Sie im Text einige Begriffe, die Ihnen im Studium bisher nicht begegnet sind. Deshalb gibt es am Ende des Buchs ein **Glossar**, in dem Sie kurze Erklärungen dieser Begriffe finden können. Im **Register**

können Sie zudem für die wichtigsten Begriffen nachschlagen, wo sie im Buch erklärt werden.

In allen Kapiteln sind **Übungsaufgaben** zu finden. Die Lösungen dazu finden Sie am Ende des Buches. Eine längere Zusatzaufgabe finden Sie unter www.narr-starter.de

Dieser STARTER behandelt quantitativ vorgehende empirische Forschung aus der rezeptiven Perspektive, also der Perspektive des Lesers der Untersuchungen. Wenn Sie eigene empirische Untersuchungen planen, empfehlen wir das dazu gehörende Studienbuch *Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht* (Albert/Marx 2016) aus der Reihe narr STUDIEN-BÜCHER.

Wir danken Frau Vanessa Weihgold vom Narr-Verlag für ihre Unterstützung beim Layout und die Beantwortung der vielen Fragen dazu sowie Frau Margarete von Legat für ihre sorgfältige Lektüre des Manuskripts. Ganz besonders danken wir Tillmann Bub, der uns zu diesem Buch überredet hat und es sorgfältig Korrektur gelesen hat.

Ettlingen/Bremen im August 2017

Ruth Albert
Nicole Marx

1. Welche Fragen können empirisch untersucht werden?

Nicht alle Fragen, die man sich als Sprachlehrforscher oder Linguist stellt, sind nur mit empirischer Forschung zu beantworten. Angenommen, Sie wollen wissen, ob im Deutschen das prädikative Adjektiv – wie etwa in den romanischen Sprachen, vgl. die Beispiele in (1) – flektiert wird, so können Sie sich sicher auf Ihr eigenes Sprachvermögen als kompetenter Sprecher des Deutschen verlassen. Sie betrachten also ein paar Sätze mit prädikativen Adjektiven wie in (2).

1. *Ce lecteur est intelligent/Cette lectrice est intelligente/Nos lecteurs sont intelligents.*
2. *Dieser Leser ist klug/Diese Leserin ist klug (*kluge)/Unsere Leser sind klug (*kluge).¹*

Dabei können Sie feststellen, dass hier das Adjektiv nicht flektiert wird. Vorsichtshalber überprüfen Sie noch, ob *klug* denn überhaupt flektierbar ist (manche Adjektive wie *lila* z. B. werden ja von vielen Sprechern des Deutschen nicht flektiert, vgl. *das lila Tuch*), aber da es für Sie eindeutig *die kluge Leserin* und *die klugen Leser* heißt und es keinerlei Anlass für die Annahme gibt, dass sich das bei irgendeinem kompetenten Sprecher des Deutschen anders verhält, ist die Frage beantwortet, ohne dass man eine repräsentative Aus-

1 In der Linguistik markiert man Sprachbeispiele, die nicht den Regeln der Sprache entsprechen (man sagt, sie sind „ungrammatisch“) mit einem Sternchen (Asterisk).

wahl von Sprechern des Deutschen hätte befragen oder endlose Textkorpora durchsuchen müssen.

Leider sind aber nicht alle Fragen in der Sprachanalyse und in der Sprachlehrforschung so einfach durch Introspektion zu lösen. In diesen Fällen wendet man empirische Verfahren an. Damit ist gemeint, dass man eine systematisch zu erfassende Erfahrung zur Grundlage macht, um die zu untersuchende Fragestellung zu lösen. Im weiteren Verlauf dieses Buchs werden unterschiedliche dafür mögliche Verfahren eingeführt: Beobachtung, Befragung, Korpusanalyse, Experiment und Test. Zudem wird ein Verfahren zur Interpretation der Ergebnisse mehrerer Studien, die Metaanalyse, eingeführt.

Bei empirischen Arbeiten wird normalerweise zwischen „qualitativ“ und „quantitativ“ vorgehenden Studien unterschieden. Qualitativ vorgehende Studien sind oft nötig, wenn über einen Gegenstandsbereich noch sehr wenig bekannt ist, sodass man erst einmal eine Wissensbasis schaffen muss, auf deren Grundlage man Hypothesen entwickeln und diese in einer quantitativ vorgehenden Studie überprüfen kann. In diesem Buch besprechen wir nur quantitativ vorgehende empirische Untersuchungen.

Quantitativ-empirische Untersuchungen sind entweder hypothesenprüfend oder hypothesenbildend. Letztere sind meist vorbereitende, explorative Arbeiten, die dazu dienen, dass man Hypothesen bilden kann. Für eine Hypothese, die man empirisch überprüfen kann, gelten gewisse Bedingungen.

- Sie muss eine Aussage sein, nicht eine Frage.
- Sie muss durch die Datenerhebung in der Realität falsifizierbar sein.

- Alle Begriffe in der Aussage müssen eindeutig definiert und operationalisierbar (das bedeutet: bestimmt und messbar) sein.

Wichtig ist, dass eine Hypothese begründet sein muss. Das bedeutet nicht, dass Autoren einen „weil“-Satz zur Hypothese hinzufügen (sollten), sondern dass eine Hypothese auf der Basis vorhergehender Forschung oder zumindest logischer Deduktion entwickelt werden muss. Als Leser müssen Sie dann nachvollziehen können, warum die Annahme gemacht wurde.

Überlegen wir folgendes Beispiel:

Weibliche Babys fangen schneller an zu sprechen als männliche Babys. (Wir gehen davon aus, dass die Autoren einen guten Grund haben, diese Annahme zu machen, und dies in ihrem Artikel bereits dargelegt haben.)

Die Hypothese genügt dem ersten Kriterium (sie ist eine Aussage). Es gibt auch weibliche und männliche Babys in der Realität, und wenn wir genug davon finden, können wir die Hypothese bestätigen oder falsifizieren (also ablehnen). Danach wird es aber etwas komplizierter, denn die Begriffe müssen deutlich sein. Was bedeutet hier *schneller*? Meinen die Autoren *früher*? Was ist *sprechen*? Gilt es schon als Sprechen, wenn das Kind in der Phase des Silbenplapperns ist? Muss eine kommunikative Absicht deutlich sein? Soll nur das erste Wort gesprochen werden (und wenn ja, ab wann ist zu bestimmen, dass [papa] z.B. kein Silbenplappern ist, sondern auf eine Person verweist)? Und überhaupt: Was für *Babys* sollen untersucht werden? Denn Kinder, die z.B. trilingual aufwachsen, beginnen oft etwas später, selbst sprachlich produktiv zu werden. Man müsste daher bestimmen, ob die Babys monolingual aufwachsen. Ebenfalls wäre wichtig, dass sie keine Auffälligkeiten (z. B. Hörstörungen) aufweisen und dass sie in einem ähnlichen sozialen Umfeld aufwachsen, weil

nachgewiesen wurde, dass dies mit der Art der Kommunikation zwischen Eltern und Kindern zusammenhängt.



Aufgabe 1: Schauen Sie folgende Hypothesen an. Welche weiteren Informationen brauchen Sie, damit diese den notwendigen Kriterien genügen?

1. Frauen sprechen am Tag mehr als Männer.
2. DaZ-Kinder, die an einem sprachlichen Förderprogramm teilnehmen, verbessern ihre Sprachkompetenzen im Vergleich zu DaZ-Kindern, die an keinem Förderprogramm teilnehmen.
3. Wer bereits Englisch als Fremdsprache gelernt hat, braucht weniger Zeit, um das gleiche Niveau im Französischen zu erreichen.

Die wichtigsten Hinweise zusammengefasst

- Wenn Sie eine empirische Untersuchung lesen, versuchen Sie zuerst, zu bestimmen, ob es sich um eine quantitative, eine qualitative oder eine mixed-methods (beide Vorgehensweisen)-Studie handelt. Das können Sie normalerweise schon im Abstract herausfinden. So können Sie sich in die Lektüre besser orientieren.
- Forschung muss nicht immer empirisch sein. Es gibt sehr viele Fragestellungen, die besser auf nichtempirische Weise zu lösen sind.
- Wenn die Autoren Hypothesen aufstellen, nehmen Sie sich einige Minuten Zeit und überlegen Sie, ob diese den Anforderungen an wissenschaftlichen Hypothesen genügen. Fehlen Ihnen als Leser wichtige Informationen über die in der Hypothese genannten Konstrukte? Können alle Begriffe operationalisiert werden, d. h. zu Kategorien gemacht werden, die man in der realen Welt messen kann?

2. Woran erkennt man eine gute empirische Untersuchung?

Wie bei allen Arbeiten gibt es bestimmte Merkmale, anhand derer Sie erkennen können, ob eine Studie den geltenden Anforderungen an eine wissenschaftliche Arbeit entspricht. Wir beginnen mit einem Beispiel:

Ein Forscher will herausfinden, ob Schüler, die mit ihren Eltern zu Hause i. d. R. eine Herkunftssprache sprechen, und Schüler, die mit ihren Eltern nur Deutsch sprechen, unterschiedliche Sprachkenntnisse in der Fremdsprache Englisch haben. Um möglichst viele Schüler zu untersuchen, geht er an eine Schule, an der fast alle Schüler andere Familiensprachen haben, und an eine andere Schule, an der fast alle Schüler nur Deutsch zu Hause sprechen. Er lässt alle Schüler der 6. Klasse dieser beiden Schulen eine Bildbeschreibung schreiben, korrigiert diese und vergleicht die Fehlerquoten der Schüler mit nur deutscher Familiensprache mit denen von Schülern mit Herkunftssprachen.

Aufgabe 1: Denken Sie zuerst darüber nach: Was ist problematisch an dieser Untersuchung?



Wir gehen im Folgenden auf einige Aspekte ein, die Ihnen helfen, solche Forschungsstudien zu bewerten.

2.1 Gütekriterien

Von einer empirischen Untersuchung erwartet man vor allem drei Dinge:

- Reliabilität, d. h. dass sie mit den gleichen Ergebnissen wiederholt werden könnte
- Validität, d. h. dass sie misst, was sie messen soll
- Objektivität, d. h. dass sie unabhängig von der Person, die die Untersuchung durchführt, auswertet oder interpretiert, stets dasselbe Ergebnis liefert.

Dies klingt nach Selbstverständlichkeiten, ist jedoch im konkreten Einzelfall nicht immer leicht zu erreichen. Worauf sollten Sie dann achten, wenn Sie empirische Studien lesen und für die eigene Arbeit bewerten sollen?

Bei der Zuverlässigkeit oder Verlässlichkeit (*Reliabilität*) geht es darum sicherzustellen, dass das Ergebnis der Untersuchung bei Gleichheit aller relevanten Bedingungen immer gleich ist. Dies ist wichtig, weil wir sicherstellen wollen, dass es sich nicht um einen Einzelfall handelt. Stellen Sie sich vor, jemand wollte beweisen, dass seine Lehrmethode besser ist als eine andere, und er lässt zwei Schülergruppen dieselben Vokabeln lernen, die eine mit seiner Methode und die andere mit der anderen. Wenn er nicht vorher überprüft hat, dass die zu lernenden Vokabeln in beiden Gruppen komplett unbekannt sind (oder in beiden Gruppen gleich viele Vokabeln gleich vielen Schülern schon bekannt), dann ist dieses Messverfahren unzuverlässig und man kann bei einer anderen Messung mit anderen Vokabeln ein anderes Ergebnis erhalten.

Die fehlende Reliabilität eines Messverfahrens oder einer Untersuchungsmethode ist in der Linguistik und Sprachlehrforschung nicht immer so offensichtlich. Am einfachsten lässt sich dies prüfen, wenn man eine Untersuchung versucht zu wiederholen: Wird z. B. bei einem Experiment nicht dasselbe Ergebnis erzielt, dann könnte es sein, dass der

Zufall die Ursache für eins der erzielten Ergebnisse war. Auch bei Befragungen z. B. sollte es nicht vorkommen, dass die Ergebnisse verschieden sind, wenn Frau X oder Herr Y die Fragen stellen. Deswegen sind Replikationsstudien wichtig – das sind Untersuchungen, die unter sehr ähnlichen Bedingungen eine andere Studie zu wiederholen versuchen, um zu sehen, ob die gleichen Ergebnisse entstehen (in der medizinischen Forschung ist dies ganz besonders wichtig). Da Sie aber im Normalfall Studien nicht replizieren werden, müssen Sie die Reliabilität durch Logik zu prüfen versuchen.

Es kann auch sein, dass ein Untersuchungsergebnis zwar perfekt wiederholbar ist, aber dass die Untersuchung gar nichts über das aussagt, worüber sie etwas aussagen will, also nicht *valide* ist. Dieses Phänomen ist das Thema vieler Scherze, sogar ein humorvoller Erfinder eines Intelligenztests soll gesagt haben: „Intelligenz ist das, was meine Tests messen.“ Auch bei linguistischen Untersuchungen kann es durchaus sein, dass man etwas anderes erforscht als das, was man erforschen will. Das kann z. B. bei ungeschickt gemachten Befragungen passieren. Die Situation, bei vorgelegten Sätzen ein Urteil über die grammatische Richtigkeit abzugeben, ist für die befragten Personen recht ungewöhnlich, wenn sie nicht Linguistik studiert haben. Wir schauen dafür ein Beispiel an:

Jemand soll Sätze nach ihrer grammatischen Korrektheit beurteilen, und ein Satz ist: *Rosenkohl ist lecker*. Grammatikalisch ist der Satz korrekt, nur für einen (uns sehr befremdlich erscheinenden) Bevölkerungsteil stimmt der Satz aber inhaltlich.

Es kann also gut sein, dass die Sätze nach inhaltlichen Kriterien oder der Gebräuchlichkeit der verwendeten Wörter als „gut“ oder „schlecht“ beurteilt werden, und somit wird etwas ganz anderes gemessen als das, was man messen wollte. Auch kann man seine Daten so ungeschickt wählen, dass ein Faktor mit hereinspielt, der dafür sorgt, dass man etwas anderes misst, als man messen wollte. Generell ist es wichtig für die Planung einer Untersuchung, dass man sich überlegt, welche Ergebnisse möglich sind und ob sie geeignet sind, die Untersuchungsfrage zu beantworten.

Gerade wenn man die Leistung von Sprachlernern beurteilen will, ist die *Objektivität* nicht immer einfach zu erreichen. Einzelne Beurteiler oder Sprachlehrpersonen werden z. B. frei geschriebene Lernertexte unterschiedlich beurteilen. Deswegen ist es wichtig, genau nachzulesen, wie Leistungen bewertet wurden. Haben die Forscher sehr genaue Vorschriften für die Beurteilung gemacht, haben sie dafür gesorgt, dass das gleiche Ergebnis unabhängig vom Bewerter (sie werden oft *Rater* genannt) erreicht wird? Haben sie vielleicht zwei oder mehr Bewerter eingesetzt und die sogenannte *Interrater-Reliabilität* gemessen, weil Unterschiede in vielen Fällen trotzdem nicht vermeidbar sind (mehr hierzu in Kapitel 5)? Je stärker der Test für die Lerner standardisiert ist, umso größer ist die Objektivität. Ein Multiple-Choice-Test, bei dem es immer nur eine richtige Antwortmöglichkeit gibt, ist zum Beispiel objektiver als ein Prüfungsgespräch über ein Thema oder ein frei geschriebener Text.



Aufgabe 2: Denken Sie wieder über das Beispiel vom Beginn des Kapitels nach. Welche Beschränkungen der drei Gütekriterien könnte es geben, und warum? Was hätte der Forscher anders machen müssen?

2.2 Geltungsbereich

Darüber hinaus muss eine Untersuchung einen angemessenen Geltungsbereich haben, d. h. dass sie über mehr als nur die konkret untersuchten Personen/Situationen Aufschluss gibt. Das untersuchte Material muss so gut ausgewählt und so umfangreich sein, dass das gefundene Ergebnis für den Bereich gilt, über den man etwas aussagen will (*Geltungsbereich*). Begreiflicherweise wird sich niemand für eine Feststellung wie „Die 17 Personen, die an meiner Untersuchung zum Vokabellernen teilgenommen haben, hatten die besten Ergebnisse mit dem Lernen anhand von Mindmaps“ interessieren. Eine Untersuchung sollte also so angelegt sein, dass ein größerer Geltungsbereich abgedeckt ist als die wenigen Personen oder Daten, die tatsächlich in der Untersuchung erforscht wurden.

Damit wären wir schon bei einem weiteren Kriterium für die Qualität einer empirischen Untersuchung, nämlich die Frage, ob die Untersuchung anhand einer sinnvoll gewählten *Stichprobe* durchgeführt wurde. So gut wie nie soll nur eine Aussage über diejenigen Informanten oder Texte gemacht werden, die tatsächlich erforscht wurden, sondern normalerweise über die Gruppe, der sie angehören (und das können z. B. „kompetente Sprecher des Deutschen“ oder „deutschsprachige Zeitungstexte“ sein). Da man niemals alle kompetenten Sprecher des Deutschen oder alle deutschsprachigen Zeitungstexte untersuchen kann, muss man eine Stichprobe aus der Grundgesamtheit, über die man Aussagen machen will, ziehen.

Hier kommt es leider häufig vor, dass die Stichproben nicht gut gewählt sind oder Gruppen, die verglichen werden, aus unterschiedlichen Gründen nicht vergleichbar sind. Das